

PREDIKSI STRUKTUR TIGA DIMENSI PROTEIN ALERGEN PANGAN DENGAN METODE HOMOLOGI MENGGUNAKAN PROGRAM SWISS-MODEL

(The Prediction of Three-Dimensional Structure from Food Allergen Protein Through Homology Method Using SWISS-MODEL Program)

Hendra Wijaya dan Fitri Hasanah

Balai Besar Industri Agro (BBIA),
Jalan Ir.H.Juanda No 11, Bogor 16122, Indonesia
e-mail: faizawijaya@gmail.com

Naskah diterima 19 Januari 2016, revisi akhir 21 Oktober 2016 dan disetujui untuk diterbitkan 26 Oktober 2016

ABSTRAK. Informasi struktur tiga dimensi protein penting untuk memahami sifat dan fungsi protein ditingkat molekular secara detil. Penentuan struktur protein dengan cara pengujian di laboratorium memerlukan tahapan dan instrumentasi dengan biaya tinggi. Struktur tiga dimensi protein dapat diprediksi secara ilmiah dengan menggunakan beberapa metode secara *in silico*. Metode yang digunakan pada penelitian ini adalah metode homologi menggunakan program SWISS-MODEL. Protein yang akan diprediksi struktur tiga dimensinya adalah protein alergen pangan. Hasil penelitian menunjukkan bahwa prediksi struktur protein alergen B7TWE7 (UniProt ID) dari *Fagus sylvatica* (beechnut) dengan metode homologi menggunakan program SWISS-MODEL menghasilkan model struktur tiga dimensi protein dengan nilai identity 72% dan $e\text{-value} = 3,4 \times 10^{-58}$. Template protein yang digunakan adalah 1e09A (PDB ID) dari *Prunus avinus* (cherry).

Kata kunci: alergen, metode homologi, struktur protein, swiss-model

ABSTRACT. Information about three-dimensional structure of protein is important to clearly understand its character and function in molecular level. Protein structure determination through laboratory testing needs highly cost of steps and instrument. Three dimensional structure of protein can be scientifically predicted using several *in silico* methods. The structure of allergen protein was predicted using SWISS-MODEL as homology method in this research. The result showed that structure prediction of allergen protein B7TWE7 (UniProt ID) from *Fagus sylvatica* (beechnut) have three-dimensional structure with identity value 72% and $e\text{-value} = 3,4 \times 10^{-58}$. This study used the 1e09A (PDB ID) of *Prunus avinus* (cherry) as a template.

Keywords: allergen, homology method, protein structure, swiss-model

1. PENDAHULUAN

Pengetahuan dan pemahaman struktur tiga dimensi protein sangat penting karena dapat memberikan informasi penting dalam memahami sifat dan fungsi biokimia protein tersebut ditingkat molekular secara detil. Pada saat sekarang ini, penentuan struktur tiga dimensi protein secara penelitian laboratorium relatif sulit karena memerlukan instrumentasi yang canggih, waktu penelitian yang lama dan

memerlukan biaya yang cukup besar. Jumlah protein yang telah ditentukan struktur tiga dimensinya secara laboratorium saat ini (data terakhir 19 Januari 2016) adalah 106.816 yang disimpan dalam *server Protein Data Bank* (PDB) (RCSB PDB, 2016). Jumlah ini tentunya sangat sedikit dibandingkan dengan data sekuen (urutan asam amino protein) protein yang telah diketahui dan disimpan dalam *server database UniProt*.

Jumlah sekuen protein yang disimpan dalam database *UniProt* saat ini (data terakhir 30 Januari 2015) adalah sebanyak 110.348.070 (Uniprot, 2016).

Penelitian dalam menentukan sekuen protein lebih cepat dibandingkan dengan penelitian menentukan struktur tiga dimensi protein mendorong munculnya pemikiran untuk memprediksi secara ilmiah struktur tiga dimensi protein dari sekuen yang ada berdasarkan pada data struktur tiga dimensi protein yang telah diketahui secara laboratorium (Berman, *et al.*, 2007). Hal ini didasarkan bahwa jumlah kelompok atau keluarga protein yang terjadi di alam relatif terbatas sehingga kesamaan struktur antara dua protein dapat disimpulkan dari kemiripan sekuen. Teknik ini kemudian dikenal luas sebagai pemodelan protein (*protein modeling*) dalam perkembangan di bidang bioinformatika (Chothia, 1992).

Struktur protein terdiri dari struktur primer, sekunder, tersier dan kuartener. Struktur primer merupakan struktur yang sederhana dengan urutan-urutan asam amino yang tersusun secara linear yang mirip seperti tatanan huruf dalam sebuah kata dan tidak terjadi percabangan rantai. Struktur sekunder adalah struktur dua dimensi protein merupakan kombinasi antara struktur primer yang linear distabilkan oleh ikatan hidrogen antara gugus =CO dan =NH di sepanjang tulang belakang polipeptida. Salah satu contoh struktur sekunder adalah α -heliks (Taylor, *et al.*, 2001). Struktur tersier dari suatu protein adalah lapisan yang tumpang tindih di atas pola struktur sekunder yang terdiri atas pemutarbalikkan tak beraturan dari ikatan antara rantai samping (gugus R) berbagai asam amino. Struktur ini merupakan konformasi tiga dimensi yang mengacu pada hubungan spesial antar struktur sekunder (Murray, *et al.*, 2009; Lehninger, *et al.*, 2004). Tiga teknik pemodelan struktur tiga dimensi protein yang dikenal saat ini diantaranya adalah metode homologi/komparatif, metode *fold recognition* dan metode *ab initio* (Zaki & Bystroff, 2008). Teknik pemodelan homologi adalah pemodelan struktur tiga dimensi protein berdasarkan pada

pensejajaran barisan (*alignment*) sekuen asam amino protein target dengan protein mirip yang telah diketahui struktur tiga dimensinya secara laboratorium (*template*) (Sali & Blundell, 1993). Tingkat kesulitan pengerjaannya dengan cara homologi secara komputasi lebih mudah dibandingkan metode *fold recognition* dan metode *ab initio*. Pengerjaan dengan metode homologi juga lebih cepat dibandingkan metode lainnya (Zaki & Bystroff, 2008). Apabila tidak ada protein *template* pada *database* maka pemodelan protein dilakukan dengan metode *ab initio*. Metode *ab initio* adalah metode yang paling sulit dan kompleks dibandingkan metode lain dan perlu waktu pengerjaan yang lama. Prinsip metode *ab initio* adalah Pemodelan struktur tiga dimensi protein berdasarkan fungsi energi (hardin & Luthey, 2002). Metode ini hanya dapat digunakan terbatas untuk protein yang relatif kecil dan akurasi yang dihasilkan juga kecil (Kopp, *et al.*, 2007). Teknik pemodelan *fold recognition* relatif lebih mudah dari *ab initio* tapi lebih sulit dari homologi (Zaki & Bystroff, 2008). Prinsip pemodelan dengan metode *fold recognition* adalah membandingkan sekuen target dengan struktur *template* di *library* sehingga menghasilkan model struktur dengan nilai fold terbaik (Hung & Samudrala, 2003). Oleh karena itu, metode pemodelan homologi adalah metode pilihan terbaik untuk membangun model struktur tiga dimensi protein secara *in silico* dalam semua kasus dengan syarat telah tersedianya struktur protein *template* (Kopp, *et al.*, 2007). Pemodelan homologi juga banyak digunakan dalam *virtual screening*, disain percobaan mutagenesis dan mempelajari efek variasi sekuen (Hillisch, *et al.*, 2004; Kopp & Schwede, 2004).

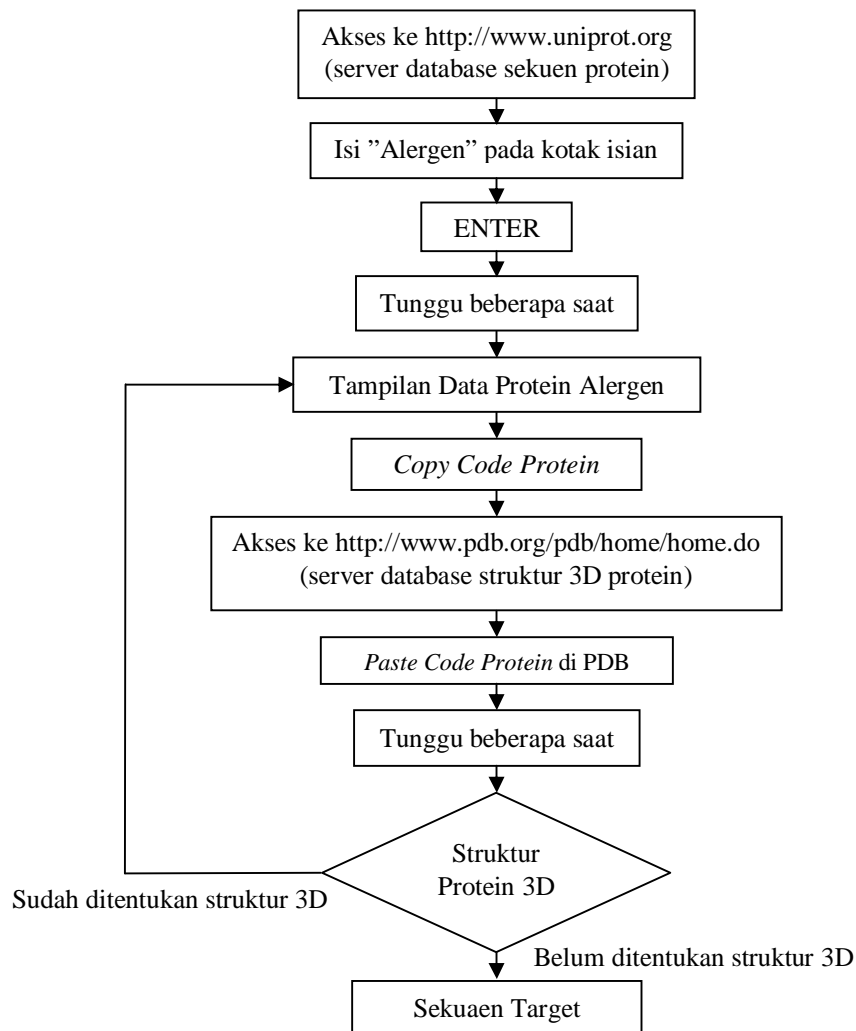
Tujuan penelitian ini adalah memprediksi atau membuat model struktur tiga dimensi protein alergen pangan menggunakan program SWISS-MODEL. Informasi tentang struktur tiga dimensi protein alergen pangan akan dapat memudahkan dalam penentuan struktur protein secara laboratorium, memudahkan dalam mempelajari mekanisme alergi

pangan oleh protein tersebut dan memudahkan dalam membuat produk pangan hipoalergenik.

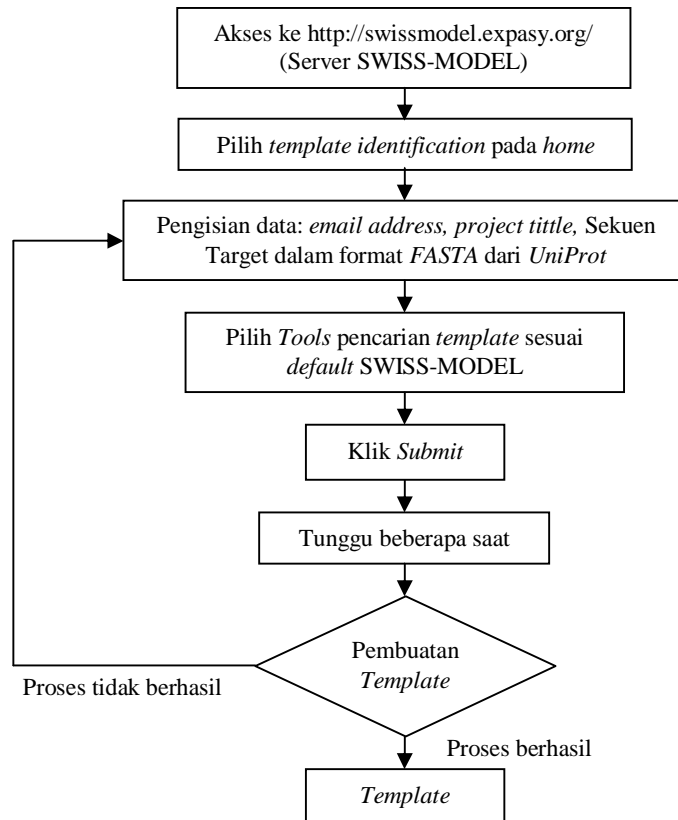
2. METODE PENELITIAN

Bahan yang digunakan dalam penelitian ini adalah laptop Acer Aspire 2930 core 2 duo t6400-2 ghz yang terkoneksi dengan internet. Koneksi internet dapat mengakses *server UniProt* di <http://www.uniprot.org>, *server Protein Data Bank (PDB)* di <http://www.pdb.org/pdb/home/home.do> dan *server SWISS-MODEL* di <http://swissmodel.expasy.org/>

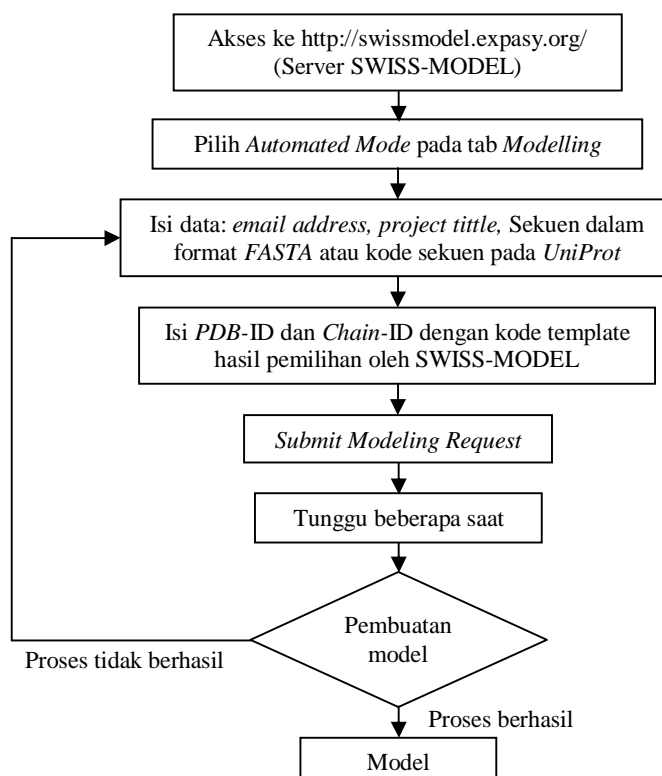
serta email individu. Tahapan penelitian yang dilakukan untuk memprediksi struktur tiga dimensi protein alergen pangan dengan cara homologi sesuai dengan protokol *Protein Structure Homology Modeling Using SWISS-MODEL Workspace* menurut Bordoli, *et al.*, (2009) dengan tahapan yang dapat dilihat pada Gambar 1 yaitu Pemilihan sekuen target protein, Gambar 2 yaitu identifikasi *template*, Gambar 3 yaitu pembuatan model dan Gambar 4 yaitu evaluasi model.



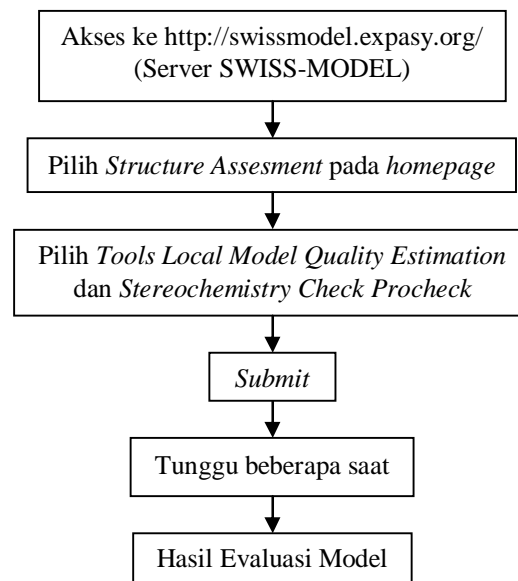
Gambar 1. Pemilihan sekuen target protein



Gambar 2. Identifikasi *Template*



Gambar 3. Pembuatan model



Gambar 4. Evaluasi model

3. HASIL DAN PEMBAHASAN

Pemilihan Sekuen Target

Protein yang dipilih adalah protein alergen pangan yang sekuen proteinnya telah ditentukan secara laboratorium tetapi struktur tiga dimensinya belum ditentukan secara laboratorium. Gambar 5 menunjukkan hasil pencarian protein alergen dalam *database UniProt*. Hasil

pencarian protein alergen adalah sebanyak 9003 protein yang diakses pada tanggal 20 Mei 2011. Berdasarkan informasi organisme pada kolom *organism* maka dipilih protein alergen dari sapi dengan kode protein pada *UniProt* Q28133 alergen untuk pengecekan di PDB yaitu di <http://www.pdb.org/pdb/home/home.do>.

Accession	Entry name	Status	Protein names	Gene names	Organism	Length
P30438	FEL1A_FELCA	★	Major allergen 1 polypeptide chain 1	CH1	Felis catus (Cat) (Felis silvestris catus)	92
P30440	FEL1B_FELCA	★	Major allergen 1 polypeptide chain 2	CH2	Felis catus (Cat) (Felis silvestris catus)	109
P19963	ALL1_OLEEU	★	Major pollen allergen		Olea europaea (Common olive)	145
P08176	PEPT1_DERPT	★	Peptidase 1	DERP1	Dermatophagoides pteronyssinus (European house dust mite)	320
Q28133	ALL2_BOVIN	★	Allergen Bos d 2		Bos taurus (Bovine)	172
P80384	ALL2_LEPDS	★	Mite group 2 allergen Lep d 2		Lepidoglyphus destructor (Fodder mite)	141

Gambar 5. Hasil pencarian protein alergen dalam *database UniProt*

The screenshot shows the PDB website interface. At the top, it says 'RCSB PDB PROTEIN DATA BANK' and 'A MEMBER OF THE PDB'. Below this, there's a search bar with 'Q28133' entered. The search results show '1 Structure'.

Gambar 6. Hasil penelusuran protein Q28133 dalam database PDB

Gambar 6 menunjukkan hasil pengecekan protein Q28133 dalam database PDB. Hasil pengecekan menunjukkan pada bagian *result count: 1 structure*. Ini menandakan bahwa protein Q28133 telah ditentukan struktur tiga dimensinya secara laboratorium dan data strukturnya telah tersimpan dalam database PDB. Jika struktur tiga dimensi protein Q28133 telah diketahui dari hasil laboratorium maka tidak perlu dilakukan prediksi terhadap protein Q28133. Prediksi struktur tiga dimensi protein hanya dilakukan pada protein yang belum ditentukan strukturnya secara laboratorium.

Untuk memudahkan dalam pencarian protein alergen yang belum ditentukan struktur tiga dimensinya secara laboratorium maka dilakukan penyaringan protein dengan memilih *identity* 50% pada *UniProt*. Gambar 7 menunjukkan hasil pencarian protein alergen dengan *identity* 50%. Hasil penelusuran menunjukkan bahwa jumlah protein alergen dalam database *UniProt* dengan *identity* 50% adalah 2.917. Protein yang dipilih adalah B7TWE7 yang selanjutnya dilakukan pengecekan kembali di PDB dengan cara yang sama pada saat pengecekan protein Q28133.

The screenshot shows the UniProt website search results. The search query is 'uniprot:(Allergen) AND identity:0.5'. The results show 2,917 results. The table below lists the results.

Accession	Status	Cluster name	Size	Members	Organisms	Length	Identity
UniRef50_A0FIJ6	★	Cluster: Stress-related protein	28	A0FIJ6 B7TWE7 B7TWE8 B7TWE6 E3W9N7 E3W9P0 E3W9P2 E3W9P1 E3W9N6 +18	Citrus sinensis (Sweet orange) Fagus sylvatica (Beechnut) Nicotiana tabacum (Common tobacco) Ricinus communis (Castor bean) Prunus dulcis x Prunus persica Castanea sativa (Sweet chestnut)	161	50%
UniRef50_A0FIV2	★	Cluster: 30 kDa salivary gland allergen-like	1	A0FIV2	Toxorhynchites amboinensis	76	50%
				A0JD32 A2NVP9 UPI0000E237E3 D37I_B4	Homo sapiens (Human) Pan troglodytes (Chimpanzee)		

Gambar 7. Hasil penelusuran protein alergen pada *UniProt* dengan *identity* 50%

Names Attributes General annotation Ontologies Sequences References Cross-refs Entry info

Customize order

Complete GO annotation...

Sequences

☐

Sequence

B7TWE7 [UniParc].

Last modified February 10, 2009. Version 1.
Checksum: 047C18AFD426A076

FASTA

160

17,453

Blast

go

102030405060

MGVFTYSESTTTVITPARLFKAFVLDADNLIPKVAPOAIRSSEIEGSGGPSTIKKITFG

708090100110120

EGSQFNMYKHRIDEIDNANFTYAYTLIEGDAISETLEKIAYEIKLVASPDGGSILKSTSR

130140150160

YHTKGDHEIKEDQIKAGKEEASGIFKAVEAYLLANPAAYH

« Hide

Gambar 8. Sequence protein B7TWE7

Hasil pengecekan menunjukkan bahwa protein B7TWE7 (kode *UniProt*) tidak terdapat di dalam *database* PDB yang berarti bahwa struktur tiga dimensinya belum ditentukan secara eksperimen. Berdasarkan hasil penelusuran di PDB, protein target untuk dibuatkan model struktur tiga dimensinya adalah B7TWE7. *Database UniProt* tersimpan informasi yang cukup lengkap tentang identitas dari protein B7TWE7 yaitu *names*, *attribute*, *general annotation*, *ontologies*, *sequences*, *references*, *cros-ref* dan *entry info*. Identitas yang penting berkaitan dengan pemodelan protein B7TWE7 adalah sekuen asam aminonya yang terdapat pada Gambar 8 masing-masing menunjukkan *name* dan *origin* serta *sequence* protein B7TWE7. Protein B7TWE7 merupakan protein alergen dalam *Fagus Sylfatica* (*Beechnut*), sejenis kacang-kacangan (Hauser, *et al.*, 2008).

Data sekuen pada Gambar 8 menunjukkan bahwa residu asam amino protein B7TWE7 berjumlah 160 asam amino dengan massa 17.453 Da. Data sekuen ini di-*release* pada tanggal 10 Februari 2009 oleh Hauser, *et al.*, (2008). Data sekuen ini akan digunakan untuk pembuatan model struktur tiga dimensinya menggunakan *server* SWISS-MODEL. Penyalinan sekuen protein B7TWE7 dari *server UniProt* ke *server* SWISS-MODEL dilakukan dengan menggunakan format FASTA.

Identifikasi Template

Tools yang digunakan untuk mengidentifikasi *template* pada *server* SWISS-MODEL adalah *InterPro Domain*

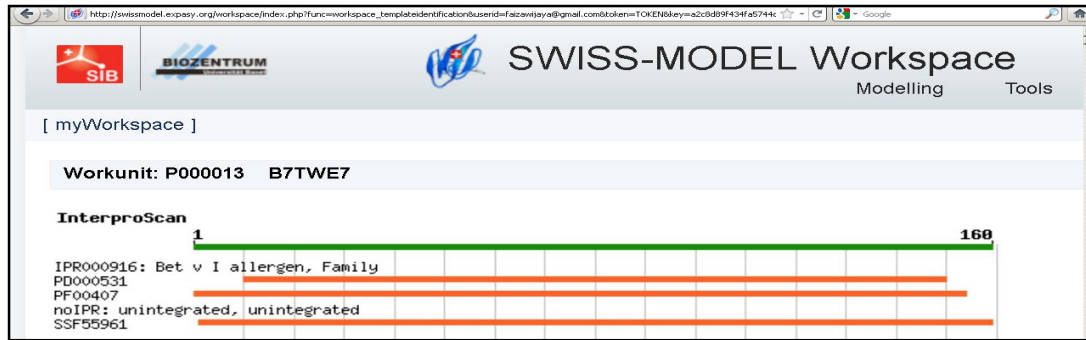
Scan, *Gapped Blast Query* dan *HHSearch Template Library Search*.

Interpro Domain Scan

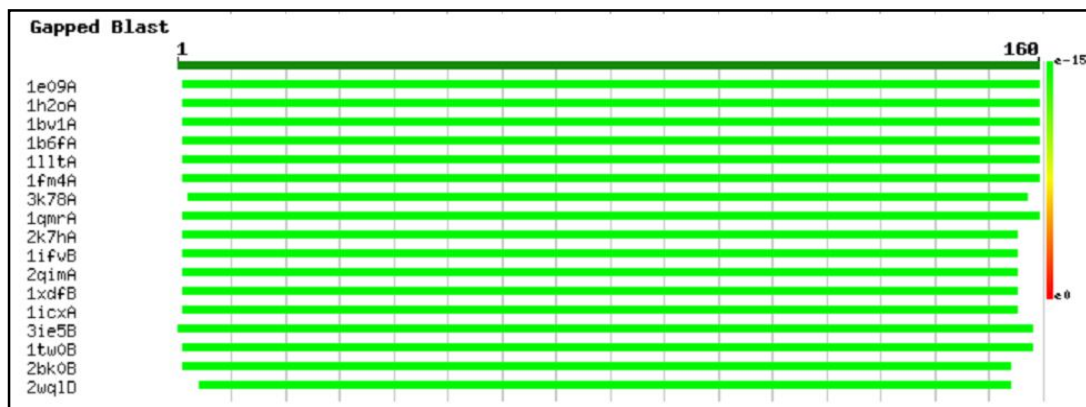
Gambar 9 menunjukkan hasil penelusuran *template* oleh *InterProscan*. *InterProscan* akan mengidentifikasi sekuen target berdasarkan domain dan sisi fungsional protein yang telah diketahui yang ditandai dengan famili spesifik. *InterProscan* terdiri dari *Pfam* yang terhubung dengan *database Pfam* yang menyimpan sekuen dan model *Markov domain protein* dan *family* protein (Mulder & Apweiler, 2007). Gambar 9 menunjukkan bahwa protein B7TWE7 dalam *database Pfam* adalah satu kelompok dengan protein PF00407 dan dalam *database InterPro* adalah dengan protein IPR000916. *InterPro* juga terkoneksi langsung dengan *database PROSITE*.

Gapped Blast Query

Gambar 10 menunjukkan hasil pencarian *template* oleh *Blast* yang diurut berdasarkan nilai *identity* dari yang besar ke yang kecil. *Green line* yang tebal menunjukkan sekuen target sedangkan *green line* yang tipis menunjukkan sekuen *template*. Terlihat bahwa sekuen *template* dimulai pada sekuen kedua sekuen target. Terdapat 16 *template* yang ditampilkan oleh *Blast* dan yang dipilih sebagai *template* adalah dengan nilai *identity* terbesar yaitu 72% dengan *e-value* (*expect value*) = 2×10^{-58} . Data hasil identifikasi *template* menunjukkan bahwa *template* dengan nilai *e-value* (221 bit) = 2×10^{-58} adalah protein alergen dari *Cherry* dengan PDB-ID 1e09A.



Gambar 9. Hasil penelusuran *template* oleh *InterProScan*



Gambar 10. Hasil pencarian *template* oleh *Blast*

HHsearch Library

Gambar 11 menunjukkan hasil pencarian *template* oleh *HHsearch library*. Penelusuran *template* oleh *HHsearch library* menghasilkan *template* 1e09A dengan *score* tertinggi 379,7. Hasil

template oleh *Blast* dan *HHsearch library* menghasilkan pilihan yang sama sehingga *template* yang akan digunakan untuk pembuatan model protein B7TWE7 adalah protein dengan kode PDB ID 1e09A (Neudecker, *et al.*, 2001).

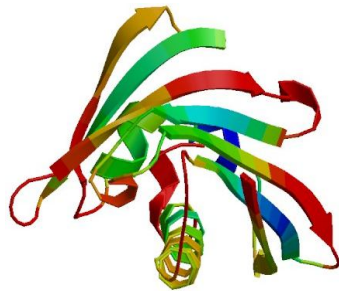


Gambar 11. Penelusuran *template* oleh *HHSearch library*

Pemodelan Sekuen Target

Setelah diketahui *template* yang akan digunakan untuk pembuatan model struktur tiga dimensi dari sekuen target maka selanjutnya dilakukan pembuatan model. *Identity* antara model dengan *template* adalah sebesar 72% sehingga untuk pembuatan model bisa menggunakan *Automated Mode*. Menurut Bordoli, *et al.*, (2009), *automated mode* bisa digunakan untuk pembuatan model apabila nilai *identity* antara sekuen target dan *template* diatas 50%. Gambar 12 menunjukkan hasil pemodelan struktur tiga dimensi protein dengan menggunakan *automated mode* pada SWISS-MODEL. Jumlah residu asam amino target adalah 160 dengan urutan 1-160 seperti pada Gambar 8.

Data *model information* menunjukkan bahwa jumlah residu asam amino model adalah 159 dengan urutan 2 sampai dengan 160. Satu residu asam amino sekuen target yaitu No 1 (Metionin, M) dihilangkan oleh SWISS-MODEL bertujuan untuk meningkatkan *identity*. *Identity* antara sekuen target dengan *template* 1e09A adalah sebesar 72,327% dengan *e-value* $3,4 \times 10^{-58}$.



Gambar 12. Hasil pemodelan struktur tiga dimensi protein B7TWE7

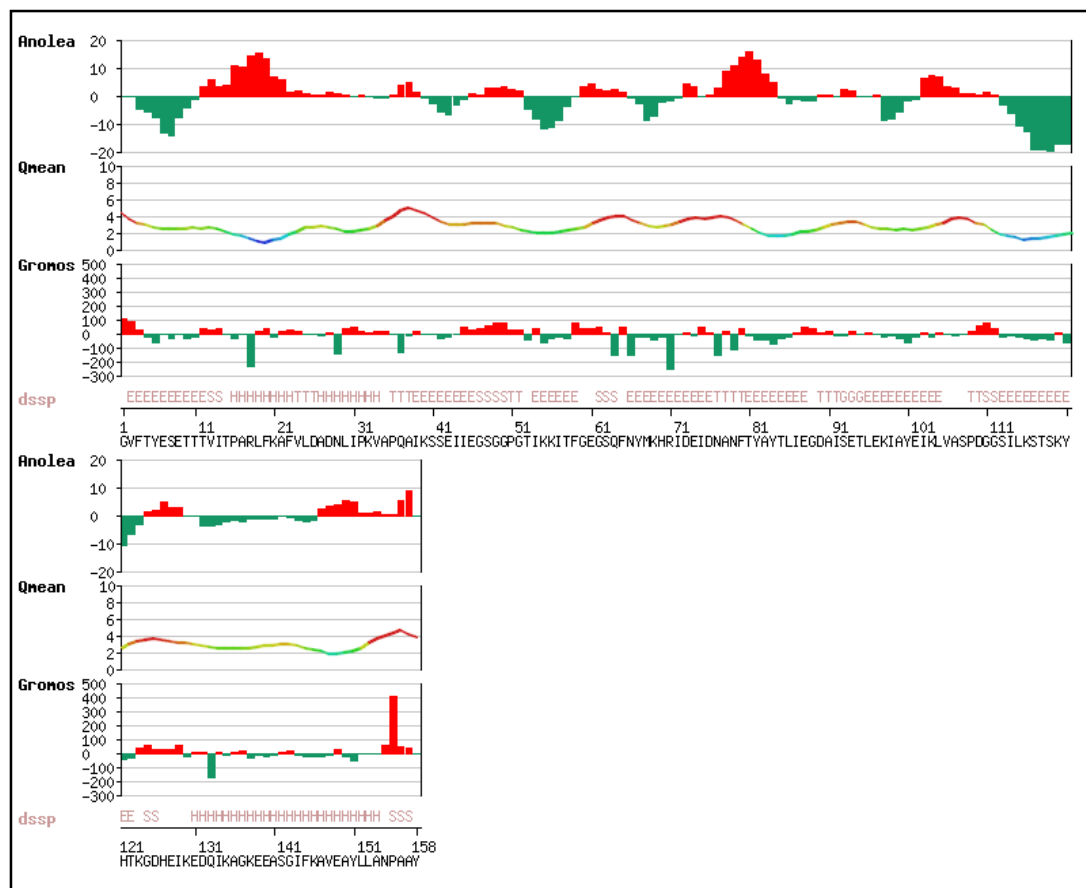
Evaluasi Model

Evaluasi struktur protein model merupakan tahap penting dalam pembuatan model (Baker & Sali, 2001). Nilai persentase identitas sekuen antara target dan *template* adalah penilaian pertama terhadap kualitas model. Semakin besar nilai % identitas antara sekuen dan target maka model yang dibuat semakin mendekati yang sebenarnya. Berdasarkan *template identification* protein alergen B7TWE7 mempunyai *identity* 72% dengan

menggunakan protein PDB-ID: 1e09 dan *Chain ID: A* sebagai *template*. Menurut Bordoli, *et al.*, (2009) akurasi model secara individu mungkin sangat bervariasi dari rata-rata mutu yang diharapkan karena berhubungan dengan tingkat homologi penjejajaran target *template*, rendahnya mutu *template*, fleksibilitas struktur dan kurang akurasi program pemodelan.

Gambar 13 menunjukkan grafik hasil Anolea, Gromos dan QMean. Anolea menunjukkan bahwa residu asam amino tersebut berada pada lingkungan yang sesuai dalam struktur proteinnya. Secara statistik, interaksi antar atom dapat digunakan untuk menganalisis kemasan (*packing*) model. Plot Anolea menunjukkan *pseudo-energy* setiap residu asam amino dalam sekuen. Nilai negatif (warna hijau) menunjukkan bahwa asam amino berada dalam lingkungan yang benar atau baik sedangkan nilai positif (warna merah) menunjukkan bahwa bagian ini tidak tersusun dengan benar. Misalnya sekuen 4-10 yaitu TYESETT dengan struktur sekunder *strained* berada dalam konformasi yang baik dengan tingkat *pseudo-energy* yang semakin rendah. Semakin besar warna merah maka residu asam amino tersebut tidak berada dalam lingkungan yang baik berdasarkan *pseudo-energy* interaksi antar atom.

Gromos merupakan perhitungan empiris yang digunakan untuk menghitung energi setiap residu di dalam model atau konformasi. Nilai negatif (warna hijau) menunjukkan bahwa asam amino berada dalam konformasi yang benar atau baik sedangkan nilai positif (warna merah) menunjukkan bahwa bagian ini tidak berada dalam konformasi benar. Penilaian energi oleh Anolea dan Gromos relatif hampir sebanding walaupun tidak semuanya sebanding karena parameter perhitungan *pseudo-energy* oleh masing-masing berbeda. Terlihat perbedaan yang cukup signifikan pada Gambar 13 pada sekuen 11-21 adalah penilaian Anolea berada pada daerah merah sedangkan penilaian Gromos tidak semuanya pada daerah merah.



Gambar 13. Grafik Anolea, QMean, dan Gromos model protein B7TWE7

QMEAN (*Qualitative Model Energy Analysis*) menggabungkan beberapa fungsi penilaian untuk memperkirakan kualitas model yang ditunjukkan pada Gambar 13. Tabel 1 menunjukkan parameter atau fungsi penilaian yang digunakan dalam perhitungan QMEAN. Dua penilaian *pseudo-energy* yang diberikan oleh QMEAN yaitu *Raw Score* dan *Z-score*.

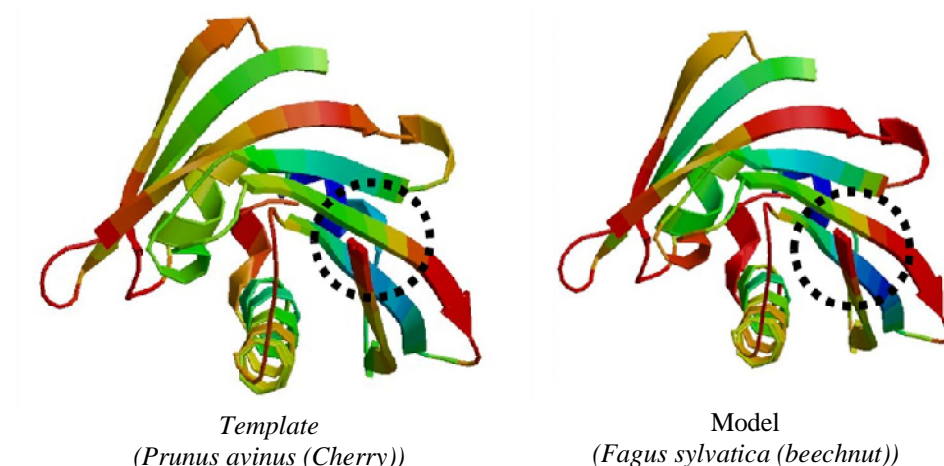
Tabel 1. Parameter yang digunakan dalam perhitungan Qmean

Parameter Penilaian	Raw Score	Z-Score
Energi interaksi C-Beta	-48,89	-1,45
Pasangan atom	-2928,65	-1,55
Energi solvasi	-11,99	-0,47
Energi sudut torsi	-18,94	-1,92
Secondary structure agreement	87,4%	0,63
Solvent accessibility agreement	75,5%	-0,47
Nilai QMEAN	0,641	-1,21

Raw score menunjukkan nilai *pseudo-energy* perhitungan secara statistik dari beberapa parameter sedangkan nilai *Z-score* diperoleh dari penentuan struktur secara eksperimen dengan ukuran yang sama dalam kristalografi X-ray. Semakin kecil nilai *pseudo-energy* pada *raw score* maka akan semakin baik sedangkan pada *Z-Score* semakin besar semakin baik.

Perbandingan Struktur Protein Hasil Pemodelan dengan Template

Gambar 14 menunjukkan perbandingan antara struktur tiga dimensi *template* dengan model. Terlihat bahwa relatif sulit untuk membedakan antara *template* dengan model. Hal ini disebabkan oleh kecilnya nilai *e-value* model yaitu $3,4 \times 10^{-58}$. Lingkaran hitam putus-putus menunjukkan salah satu lokasi perbedaan antara *template* dengan model. Pada bagian model, strukturnya berupa α -helix sedangkan pada *template* strukturnya β -sheet.



Gambar 14. Perbandingan antara struktur tiga dimensi *template* dengan model

4. KESIMPULAN

Pemodelan protein alergen B7TWE7 (UniProt ID) dari *Fagus sylvatica* (beechnut) dengan metode homologi menggunakan program SWISS-MODEL menghasilkan model struktur tiga dimensi dengan nilai *identity* 72% dan *e-value* = $3,4 \times 10^{-58}$. *Template* yang digunakan adalah 1e09A (PDB ID) dari *Prunus avinus* (cherry). Saran penelitian adalah perlu dilakukan penentuan struktur protein alergen pangan secara laboratorium untuk memastikan struktur protein sesungguhnya dan membandingkannya dengan hasil prediksi atau pemodelan.

DAFTAR PUSTAKA

- Baker, D. & Sali, A. (2001). Protein structure prediction and structural genomics. *Science (New York, N.Y.)*, 294(5540), 93–6. doi:10.1126/science.1065659.
- Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. (2007). The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Research*, 35(Database issue), D301–3. doi:10.1093/nar/gkl971.
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, 4(1), 1–13. doi:10.1038/nprot.2008.197.
- Chothia, C. (1992). Proteins. One thousand families for the molecular biologist. *Nature*, 357(6379), 543–4. doi:10.1038/357543a0.
- Hardin, C., Pogorelov, T.V. & Luthey-Schulten, Z. (2002). Ab initio protein structure prediction. *Current Opinion in Structural Biology*, 12(2), 176–81. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11959494>.
- Hauser, M., Wallner, M., Erler, A., Briza P. & Ferreira, F. (2008). *Cloning and characterization of Fags 1, a Bet v 1 homologous pollen allergen from beech*. EMBL/GenBank/DBJ databases.
- Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today*, 9(15), 659–69. doi:10.1016/S1359-6446(04)03196-4.
- Hung, L-H. & Samudrala, R. (2003). PROTINFO: Secondary and tertiary protein structure prediction. *Nucleic Acids Research*, 3(13), 3296–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12824311>.
- Kopp, J. & Schwede, T. (2004). Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, 5(4), 405–16. doi:10.1517/14622416.5.4.405.
- Kopp, J., Bordoli, L., Battey, J. N. D., Kiefer, F. & Schwede, T. (2007). Assessment of CASP7 predictions for template-based modeling targets. *Proteins*, 69 Suppl., 8, 38–46. doi:10.1002/prot.21753.
- Mulder, N. & Apweiler, R. (2007). InterPro and InterProScan. *Comparative*

- Genomics*, 396, 59-70. doi:10.1007/978-1-59745-515-2 5.
- Murray (2006). *Harper's illustrated biochemistry*. Boston, Mass.: McGraw-Hill.
- Nelson, D.L., Cox, M.M. & Lahn timer, A.L. (2005). *Lehninger: Principles of biochemistry*. New York: W.H. Freeman and Company.
- Neudecker, P., Schweimer, K., Nerkamp, J., Scheurer, S., Vieths, S., Sticht, H. & Rös ch, P. (2001). Allergic cross-reactivity made visible: solution structure of the major cherry allergen Pru av 1. *The Journal of Biological Chemistry*, 276(25), 22756-63. doi:10.1074/jbc.M101657200.
- RCSB Protein Data Bank (RCSB PDB) Web site. (2016). *PDB Current Holdings Breakdown*. Retrieved January 19, 2016, from <http://www.rcsb.org/pdb/statistics/holdings.do>.
- Rouvinen, J., Rautiainen, J., Virtanen, T., Zeiler, T., Kauppinen, J., Taivainen, A. & Mäntyjärvi, R. (1999). Probing the molecular basis of allergy. three-dimensional structure of the bovine lipocalin allergen Bos d 2. *The Journal of Biological Chemistry*, 274(4), 2337-43. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9891000>.
- Sali, A. & Blundell, Tzaki. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779-815. doi:10.1006/jmbi.1993.1626.
- Taylor, W.R., Alex, C.W., May, Nigel, P. Brown y & Andras Aszodi. (2001). *Protein Structure: Geometry, Topology and Classification*. London: The Ridgeway.
- Uniprot Web site. (2016). *Sequence archive*. Retrieved January 19, 2016, from <http://www.uniprot.org/uniparc/>.
- Zaki, M. J. & Bystroff, C. (2008). *Protein Structure Prediction*, 2nd ed. USA: Humana Press/Springer.